

더핑크퐁컴퍼니 애플리케이션 구독 추적 시스템 구축 이벤트 데이터 처리 파이프라인 개선



더핑크퐁컴퍼니는 기업 비즈니스 성장 과정에서 데이터가 폭증하는 경험에 직면했다. 모바일 애플리케이션을 비롯해 유튜브 등 다양한 환경에서 콘텐츠를 서비스하며 데이터 소스가 다양해졌다. 한 번에 처리해야 할 데이터 양이 많아지며 데이터 처리 파이프라인을 개선하기로 한 더핑크퐁컴퍼니는 클라우드 기반의 분석용 데이터웨어하우스(DW) 서비스를 찾던 중 스노우플레이크를 접하게 됐다.

글로벌 패밀리 엔터테인먼트 기업 더핑크퐁컴퍼니는 유튜브 영상 누적 조회수 850억뷰를 돌파하는 등 콘텐츠를 통해 세계에 즐거움을 전하고 있다. 특히 ‘핑크퐁 아기상어’ 콘텐츠는 25개국 언어로 글로벌 서비스되며 한국의 대표적인 지식재산권(IP)으로 자리 잡았다. 더핑크퐁컴퍼니는 스노우플레이크를 DW로 활용해 중앙화한 데이터 아키텍처를 구현하고 있다.

기존의 복잡한 데이터 아키텍처

사업 초기 인터넷데이터센터(IDC)의 레거시 온프레미스 환경을 활용하던 더핑크퐁컴퍼니는 사업 확장과 함께 아마존웹서비스(AWS)를 기반으로 클라우드 환경을 구축했다. 이후 대용량 데이터를 스노우플레이크에 올렸다.

기존에 구축해뒀던 데이터 아키텍처는 데이터가 여러 데이터베이스에 나뉘어져 있어 데이터베이스 각각의 특성을 고려하면서 분석 업무를 수행해야 했다. 이로 인해 전체적인 아키텍처 구조가 복잡했고 관리해야 할 포인트도 많았다. 또한 서비스마다 데이터가 일부 중복되

는 문제가 발생하거나 데이터의 일관성이 떨어지는 문제도 있었다.

아울러 다양한 플랫폼에서 나오는 데이터와 앱 데이터를 통합해 분석하고자 했던 더핑크퐁컴퍼니는 엘라스틱서치(ElasticSearch), 로그스태시(Logstash), 키바나(Kibana)로 구성된 ELK 환경에 데이터를 적재해 앱 이벤트와 함께 통합 분석을 수행하고 있었다. 하지만 이 또한 별도의 데이터 준비를 위한 서버를 구축하거나 대시보드를 추가로 개발해야 하는 등 오히려 신경써야 할 요소가 많았다. 또한 유튜브와 같은 데이터는 구글 클라우드 ‘빅쿼리(BigQuery)’에 적재한 뒤 분석을 시도했는데, 빅쿼리의 경우 원하는 이벤트만 선택해 적재하는 작업이 어려웠고 원하는 이벤트만 남긴 후 나머지 데이터를 지워야 하는 추가 프로세스 또한 발생했다. 결과적으로는 AWS 환경과 구글클라우드 플랫폼 간의 데이터 통합 이슈도 생겼다.

이와 같은 문제를 해결하기 위해 더핑크퐁컴퍼니는 메인 데이터 아키텍처를 스노우플레이크 기반으로 구축했다. 스노우플레이크로 중앙화된 데이터 아키텍처를 구성해 분석 업무의 효율을 높이고, 데이

“ 데이터를 조금 더 면밀히 확인하고 이를 바탕으로 좋은 의사결정을 내리는 것은 성장하는 회사가 가져야 할 덕목입니다. 스노우플레이크를 활용하면 더 세밀한 데이터 분석과 더 빠른 의사결정이 가능합니다. - 손동우 더핑크퐁컴퍼니 CTO

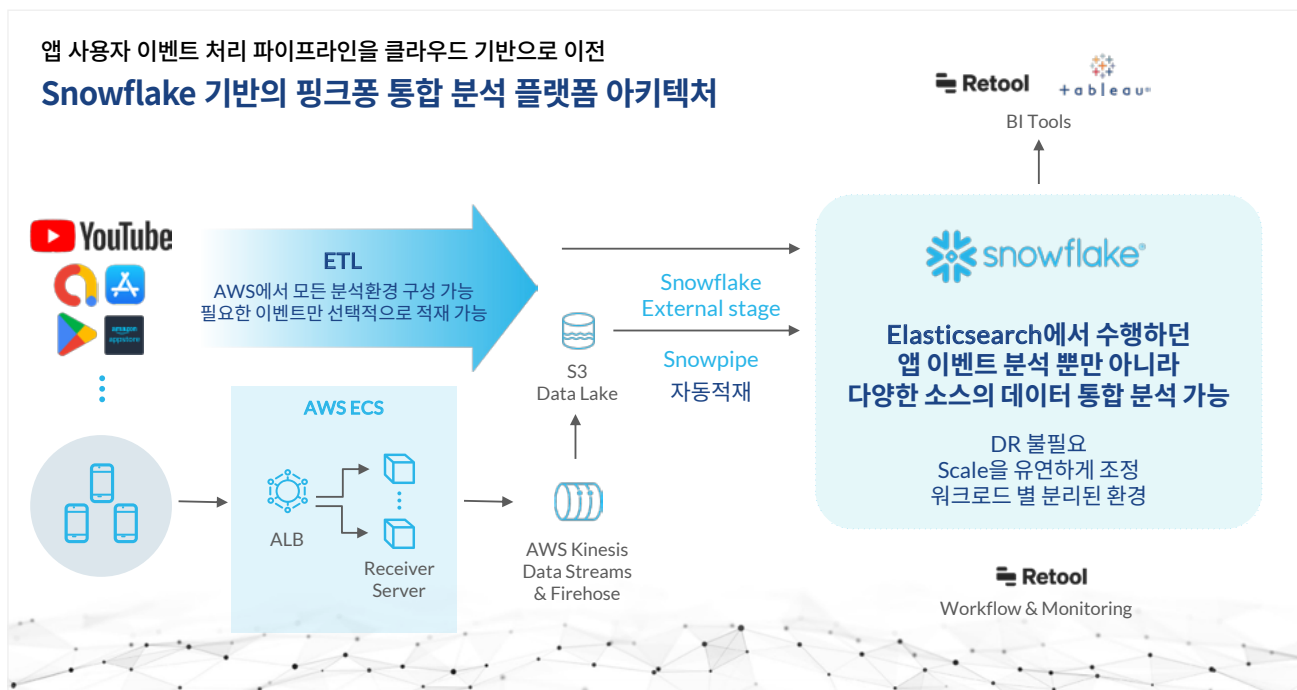
터 관리 포인트를 줄이기 위한 결정이었다. 개발 환경 통합이 편리하고 사용 패턴에 따른 스케일을 유연하게 조정할 수 있는 장점도 스노우플레이크를 선택한 이유다. 더핑크퐁컴퍼니는 특히 사용자의 이벤트성 데이터 처리 파이프라인을 개선하는데 스노우플레이크를 활용해 업무 효율 향상을 꾀했다.

도입 과정과 성과

초기 도입 과정에서 스노우플레이크코리아의 도움을 받아 PoC를 수행했던 더핑크퐁컴퍼니는 크기와 클러스터 수 조정이 용이한 스노우플레이크의 장점을 살려 데이터 처리에 사용하는 DW와 대시보드에서 데이터를 조회할 때 사용하는 DW를 나눠 구축했다. 가시성이 높은 스노우사이트(Snowsight) 인터페이스도 심분 활용했다. 한예성 더핑크퐁컴퍼니 SW 엔지니어는 “쿼리의 히스토리나 비용 모니터링 등이 가능하고 데이터 권한 제어나 공유도 쉬웠다”고 밝혔다. 또 스노우플레이크는 SQL을 비롯해 파이썬(Python), 노드닷제이에스(Node.js) 등의 다양한 개발 언어를 지원하기 때문에 개발자들의 적응도 쉬웠다는 설명이다.

또한 더핑크퐁컴퍼니는 앱에서 발생하는 사용자 이벤트와 각종 플랫폼에서 나오는 매출 및 구독 데이터를 통합, 모든 예상 매출을 분석하는 시스템 구축에 성공했다. 새로 출시한 ‘아기상어 퀴즈월드’ 앱 사용자가 어떤 경로를 통해 구독까지 완료했는지 추적하는 시스템도 구축했다.

더핑크퐁컴퍼니는 스노우플레이크를 통해 또 다른 분석 시스템도 구축할 계획이다. 향후 머신러닝(ML) 분석에 스노우파크(Snowpark)를 활용해 콘텐츠 추천 시스템을 구동하는 한편, 여러 서비스에 분산된 데이터 워크플로우 또한 통합할 방침이다. 아울러 스트림릿(Streamlit)을 통해 대시보드 개발 환경까지 통합함으로써 데이터 분석에 필요한 모든 업무를 스노우플레이크 상에서 수행할 계획까지도 가지고 있다.



스노우플레이크에 대하여

스노우플레이크는 데이터 클라우드를 통해 데이터를 가치있게 활용할 수 있도록 합니다. 데이터 사일로 없이 통합된 데이터를 활용할 수 있고, 데이터를 안전하게 공유하며, AI/ML 및 다양한 분석 워크로드를 지원합니다. 2023년 7월 30일 기준, Forbes Global 2000 기업 중 639개 기업을 비롯한 다양한 산업 분야의 8,537개사의 고객이 스노우플레이크 데이터 클라우드를 사용하여 비즈니스를 강화하고 있습니다. 자세한 내용은 www.snowflake.com/ko에서 확인할 수 있습니다.